Erik Ter-Gabrielyan
STAT100 Section 0122

Problem Set 3

1a.

```
> load("C:/Users/ErikTG/Downloads/Course_Data_Set.RData")
> # Erik Ter-Gabrielyan
> # Problem Set 2
> summary(Course_Data_Set$Height_inches)
   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.    NA's
  56.00   63.00   66.00   66.23   70.00   82.00       9
```

1b.

```
> # Erik Ter-Gabrielyan
> summary(Course_Data_Set$HS_GPA)
   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.    NA's
  0.000   3.000   3.425   3.331   3.723   4.000      27
```

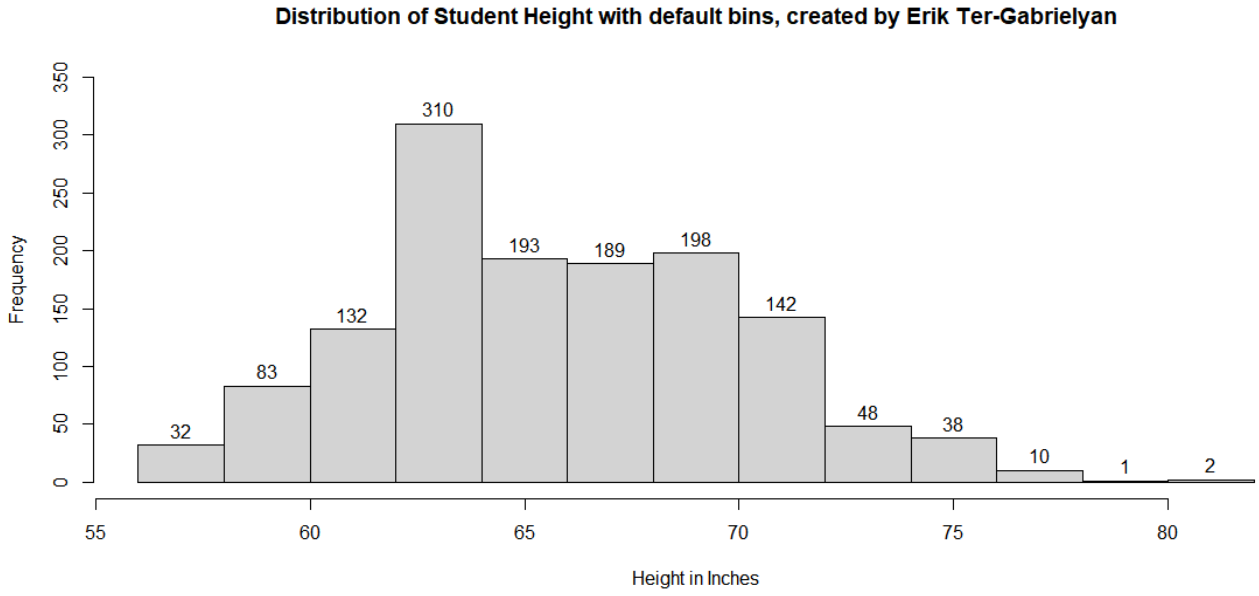
1c.

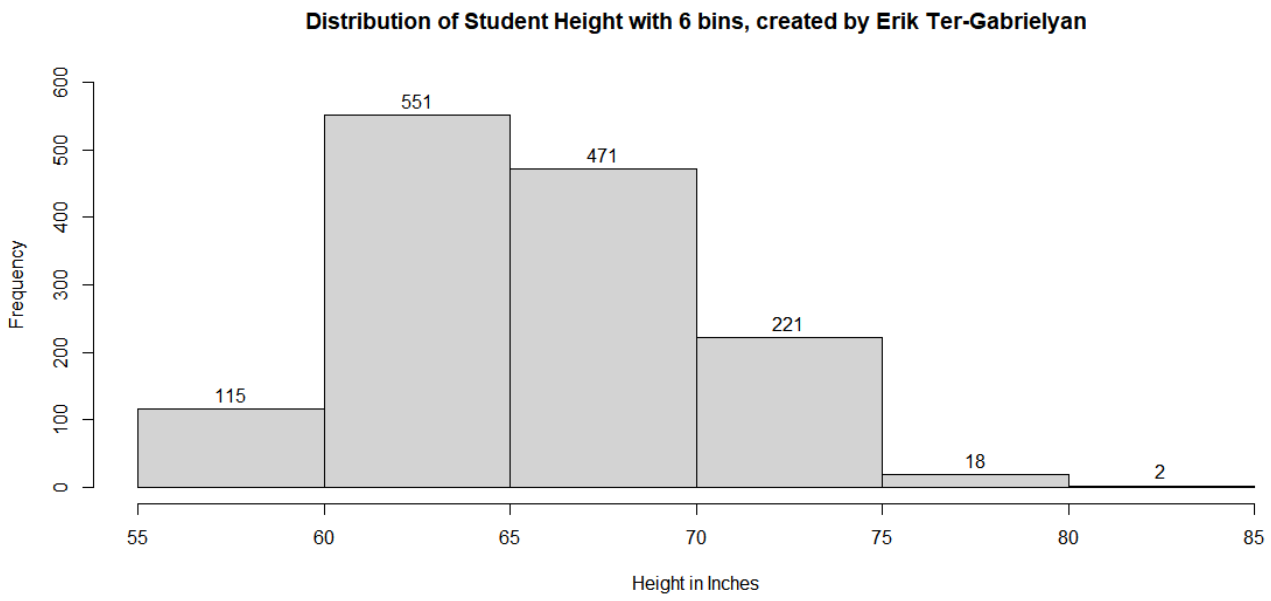
```
> 1.5*(70-63)
[1] 10.5
> # Since the Q3+(1.5 * IQR) of the Height_inches variable is 80.5, and the value given is 80, 80 would not be considered an outlier since it is less than 80.5
```

1d.

```
> 1.5*(3.723-3)
[1] 1.0845
> 3-1.0845
[1] 1.9155
> # Since the Q1-(1.5 * IQR) of the HS_GPA variable is 1.9155, and the value given is 1.85, 1.85 would be considered an outlier since it is less than 1.9155
```

2a.





> hist(Course_Data_Set$Height_inches,main="Distribution of Student Height with default bins, created by Erik Ter-Gabrielyan",ylim=c(0,350),xlab="Height in Inches",labels=TRUE)
> hist(Course_Data_Set$Height_inches,main="Distribution of Student Height with 6 bins, created by Erik Ter-Gabrielyan",xlab="Height in Inches",ylim=c(0,600),breaks = 6,labels=TRUE)

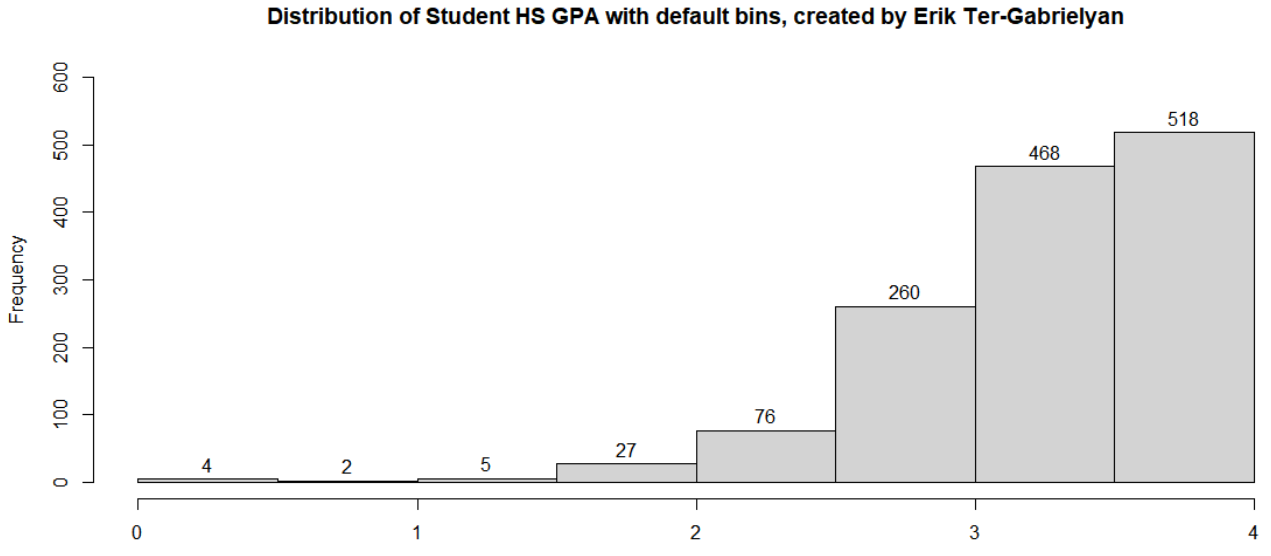
2b.

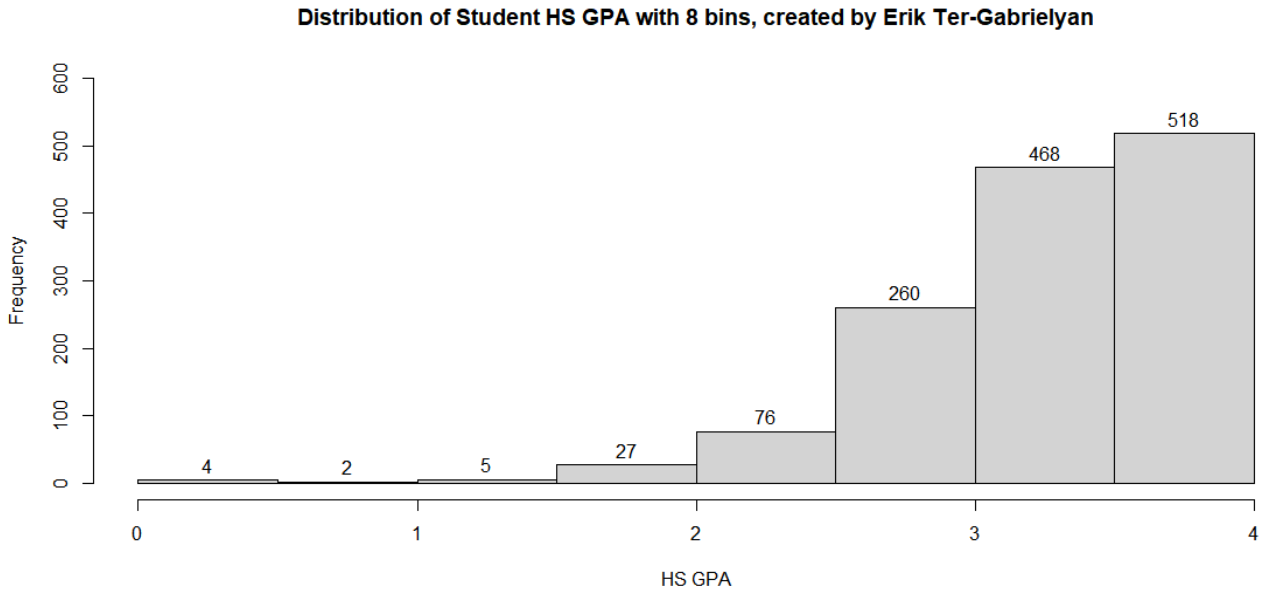
The distribution of Student Height in Inches is positively skewed, with a center shifted left between 60 and 70 inches with 2 outliers above 80.5. Although not entirely different, the distribution with 6 bins does not show as much detail in the shape as the one with more bins.

2c.

I used the 6 bins distribution, and tallied up all the bars. The bars equalled 1378. I then divided that by 2 to get 689, which is where the median lies. Then, I began counting as many bars as I could until I reached or passed 689. In this case, I counted 115 + 551, which added up to 666, and the next bar was 471. This meant that the median was in the beginning of the 65-70 category.

2d.





> hist(Course_Data_Set$HS_GPA,main="Distribution of Student HS GPA with default bins, created by Erik Ter-Gabrielyan",ylim=c(0,600),xlab="HS GPA",labels=TRUE)
> hist(Course_Data_Set$HS_GPA,main="Distribution of Student HS GPA with 8 bins, created by Erik Ter-Gabrielyan",xlab="HS GPA",ylim=c(0,600),breaks = 8,labels=TRUE)

2e.

The distribution is negatively skewed with its center shifted far right, with a min of 0 and max of 4 and outliers below 1.85, making there 11 total outliers. The histograms are exactly the same regardless of the breaks/bins I used, so there was no difference in which I could use to gather the information.

2f.

Same as the method for 2c, I added all the bars to find the total, 1360, divided by 2 to find the median place which is 680. Then I added bars until I reached or would pass 680, in this case 4+2+5+27+76+260+468 = 842, so that means the median has to occur in the 468 bar, which is between 3 and 3.5 HS GPA.